The first author, then a junior student, decided not to pursue her academic career at least, not immediately after her graduation, so we are not writing up a proceedings paper.

# Can Japanese speakers compensate for coarticulation due to [l] and [r]?

FAJL 8 @ Mie University
Chika Takahashi & Shigeto Kawahara
(Keio University)
2016/2/18

Contact: kawahara@icl.keio.ac.jp

# Introduction

# Context effect in speech perception

* Speech production of a segment is influenced by surrounding segments (a.k.a. coarticulation).

* Speech perception of a segment is likewise influenced by surrounding segments.

* Classic work by Ladefoged & Boardbend (1957), which shows that the perception of vowel height is affected by the precursor sentence.

# Ladefoged & Boardbend (1957)

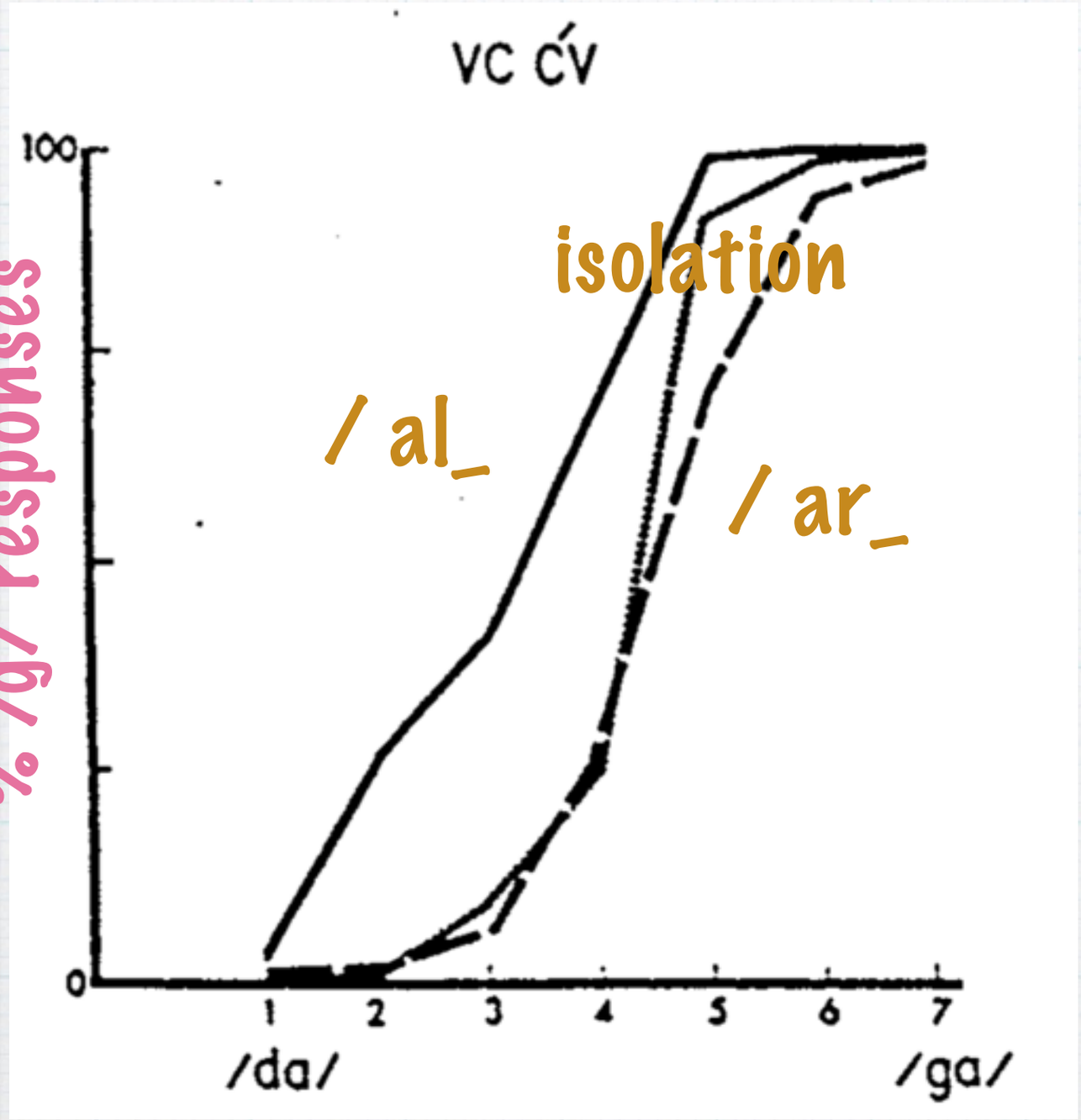
* [bɪt]
* [bɛt]

# Context effect as normalization

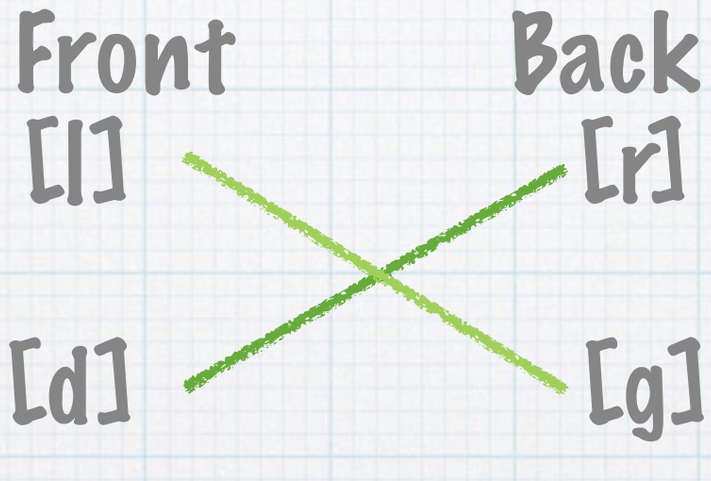
* Context effect is a way to deal with context-dependent variability (due to coarticulation).

* Mann (1980): Given a [d]-[g] continuum, English listeners hear more of the continuum as [g] after [l] than after [r].

# Mann (1980)



# Compensation for coarticulation



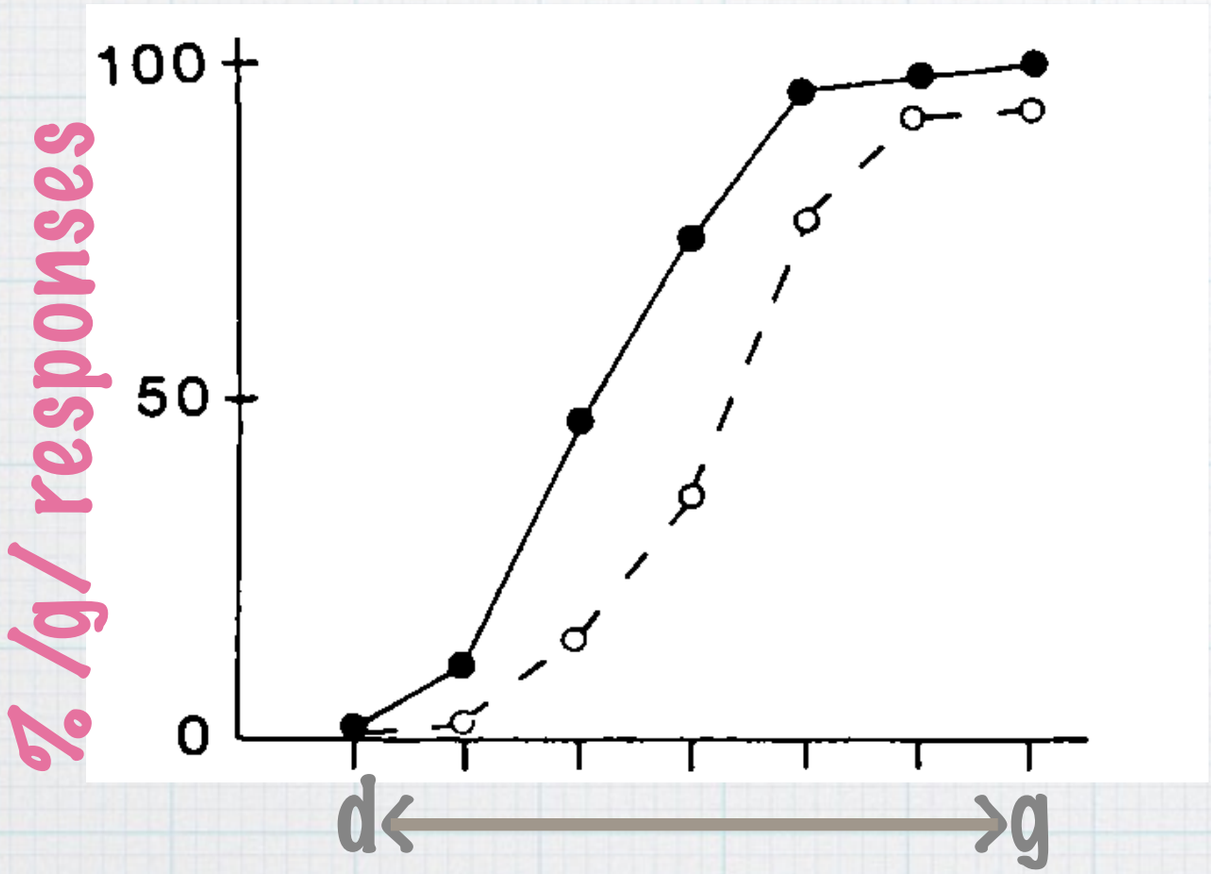

* Listeners assume that after [l], the speaker's tongue position is fronted, make up this assumed fronting, and are more likely to judge the continuum as [g].
* This theory is called "compensation for coarticulation" (Fowler 2006, P&P).

# Mann (1986)

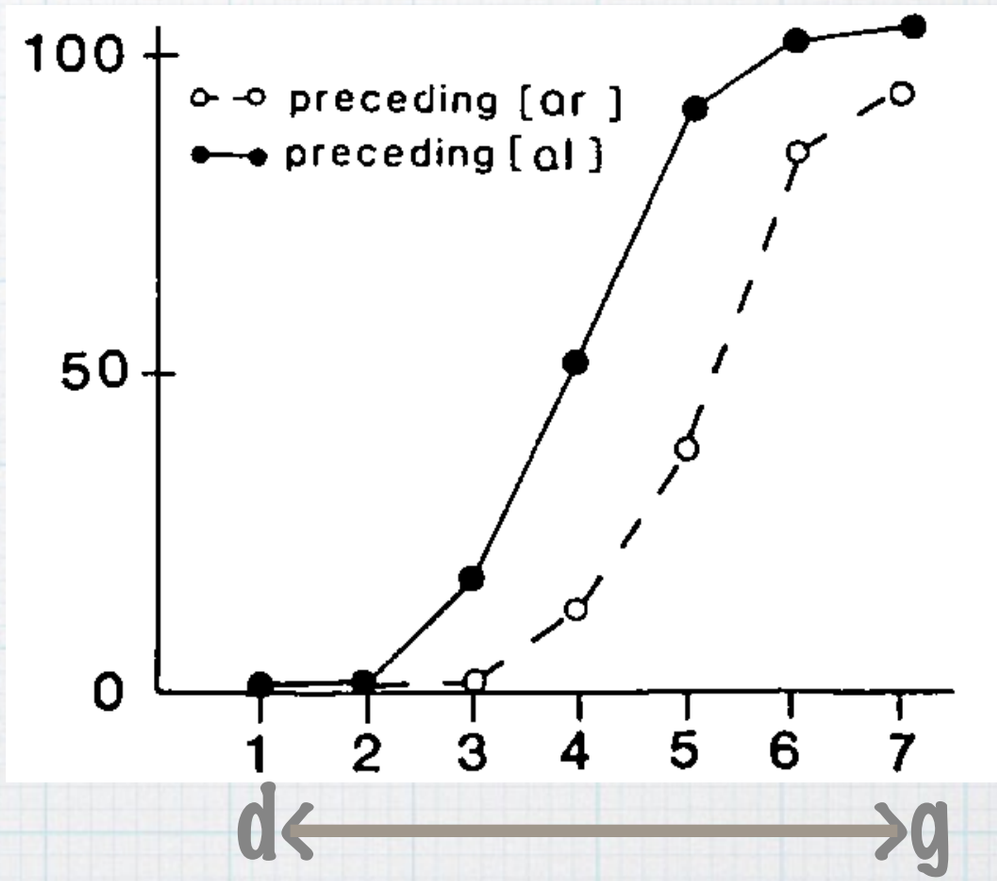
* Japanese listeners are famously unable to hear the difference between English [l] and [r] (Goto 1971, et seq.).

* Mann (1986) argues that Japanese speakers cannot hear this difference, but they nevertheless compensate for coarticulation due to [l] and [r].

# Mann (1986)



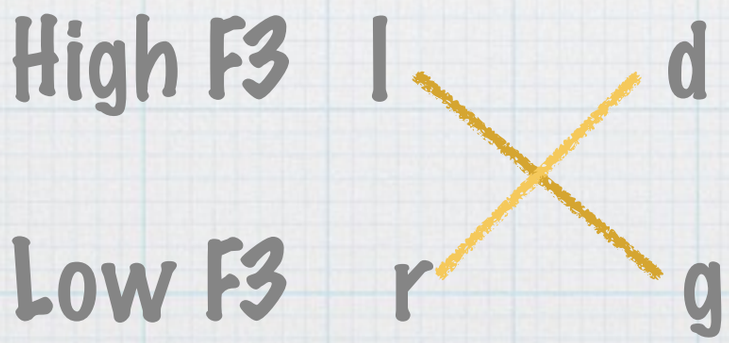

English listeners



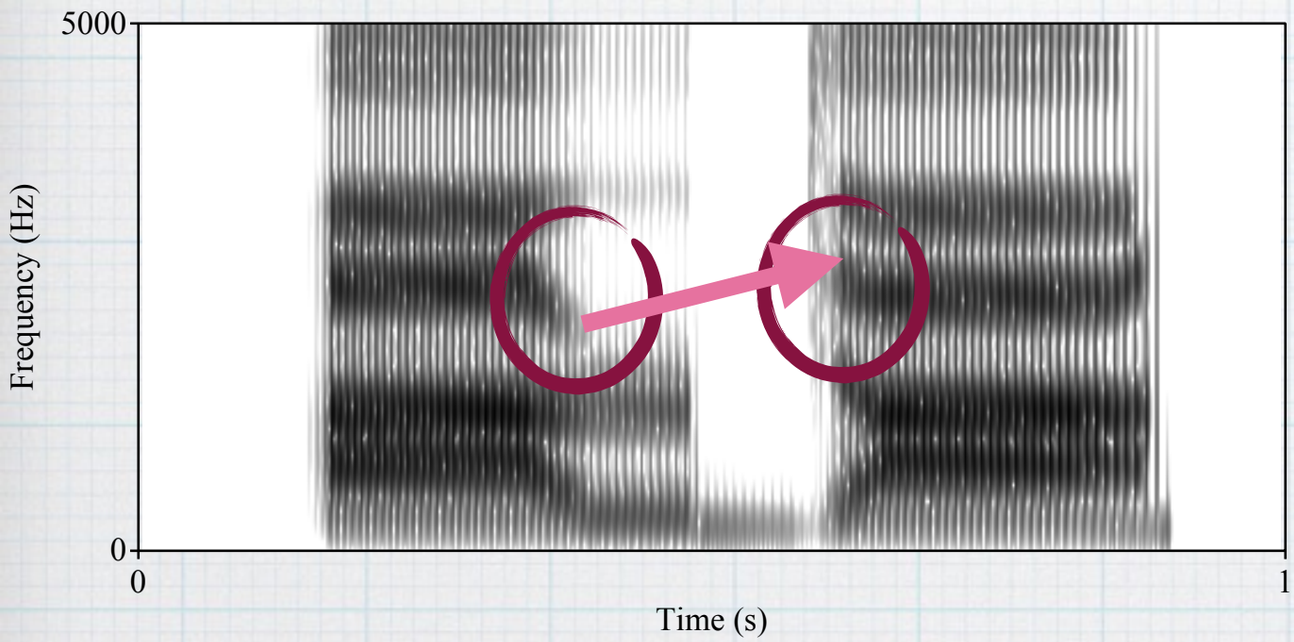

Japanese listeners

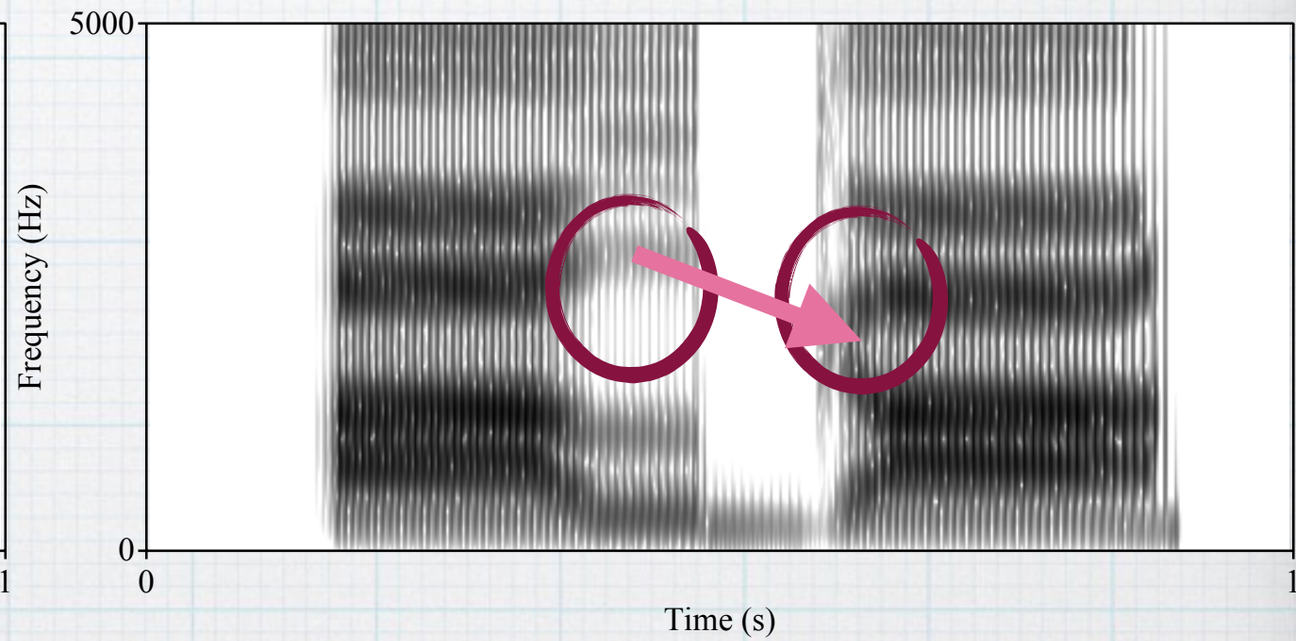
# General auditory contrast?

* But why? Japanese listeners are aware of the different articulatory gestures of [l] and [r] anyway?
* Mann (1986) attributes this result to a universal perceptual mechanism.
* An alternative explanation of contrast effect is general auditory contrast (Kluender & Lotto 1998)



# Low after high
# High after low



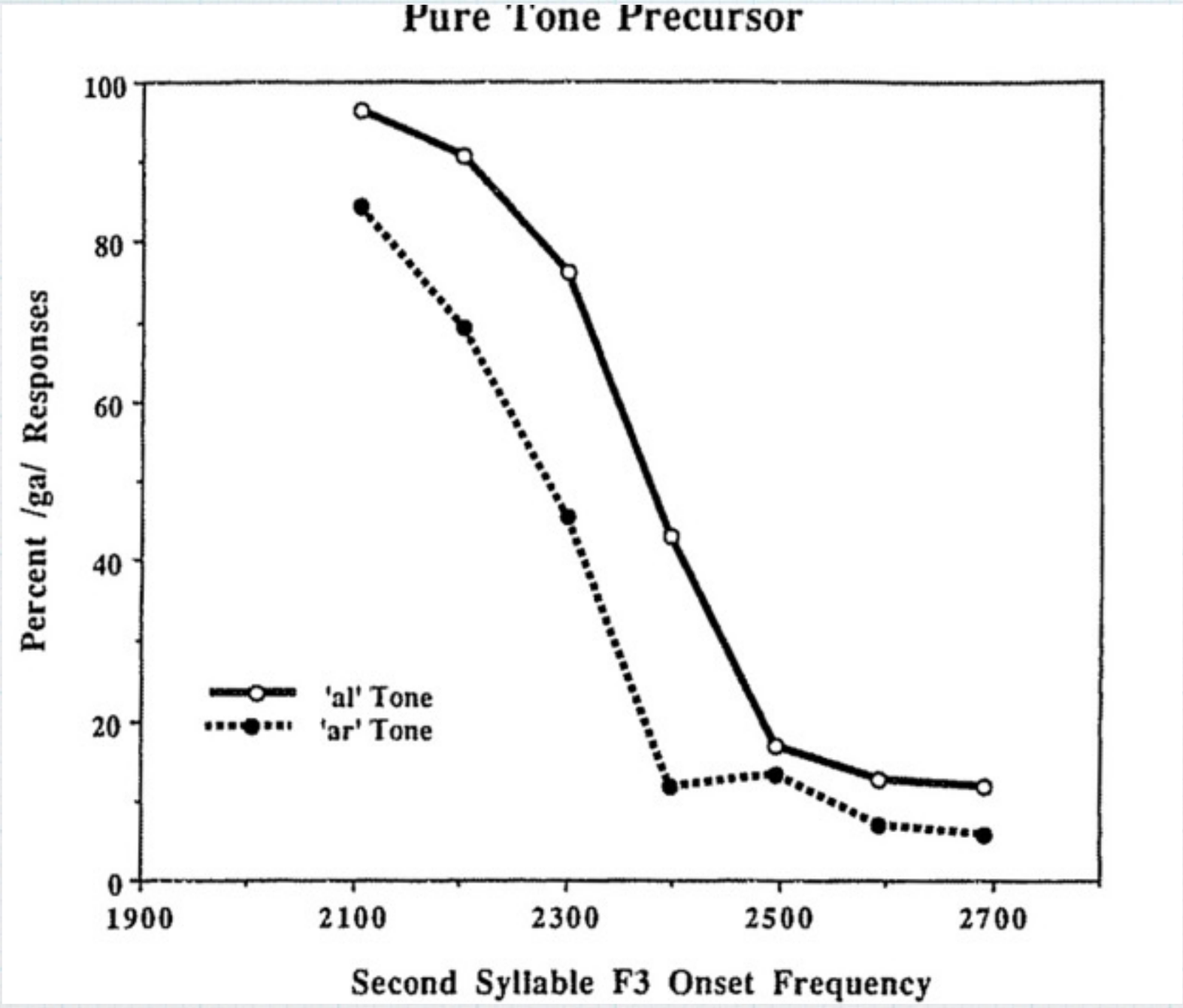
# General auditory contrast

* In this theory, listeners do not need to know how [l] and [r] are articulated.

* Context effect arise as the result of auditory contrast.

* This theory is further supported by the observation that non-speech precursors can cause context effect (Lotto and Kluender 1998).

# Lotto and Kluender's (1998) results




Though see Viswanathan et al. (2009, 2012) for a reply.

# The Current Experiment

# Questions about Mann (1986)

* Do all Japanese speakers show contrast effect due to [l] and [r]?
    * The general auditory contrast theory predicts that they should.
* Does the magnitude of contrast effect correlate with the ability to distinguish [l] and [r]?
    * The compensation for coarticulation theory (perhaps) predicts a positive correlation.
* Is context effect universal (cf. Beddor et al. 2002; Kang et al. in press; Yu et al. 2013)?
* In Mann (1986), context=natural speech; target=synthetic speech. There could have been some unnaturalness.

# The current experiment

* The current experiment tested the ability to distinguish [l] and [r], and the effect of context effect due to [l]-[r] at the same time, from the same participants.
* The current experiment also used a synthetic [l]-[r] continuum (Kingston et al. 2014 et al.)
    * The auditory contrast theory predicts that the higher the F3 is, the more [g] response we should get.
* Would we observe a simple linear increasing effect of the [l]-[r] continuum on the [d]-[g] judgment?

# Stimulus structure

* [aXYa] where X is a {r-l} continuum and Y is a {d-g} continuum.

[l]-[r] continuum before [d]

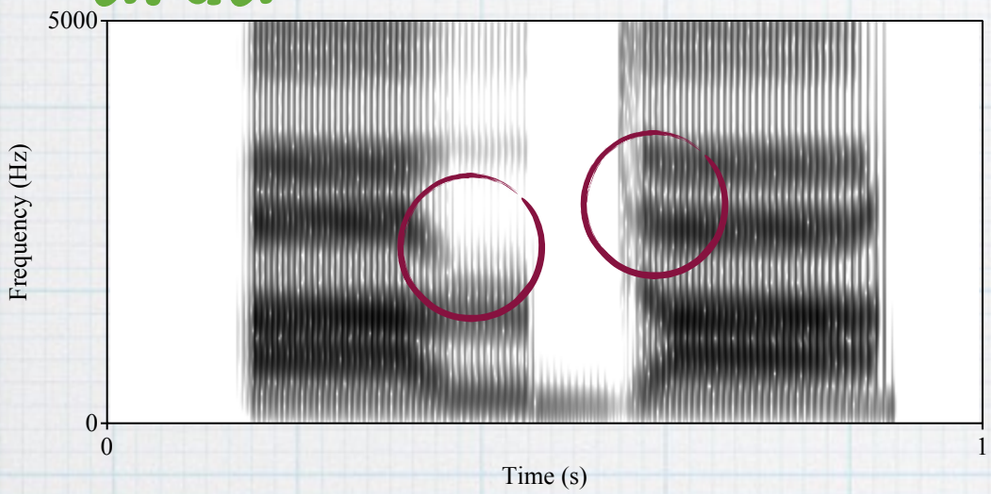
[d]-[g] continuum after [l]

# Method: Stimuli

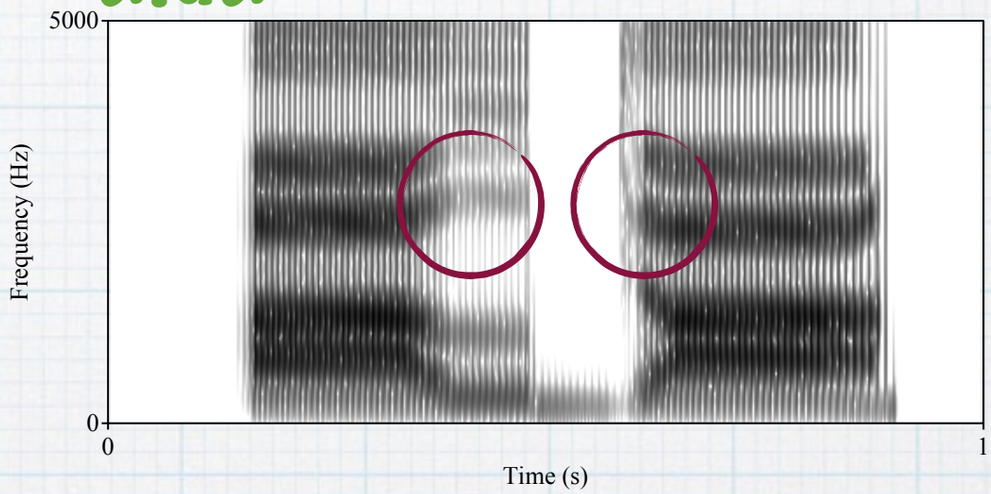
- The two surrounding vowels are always identical, [a] with F3 of 2500 Hz.
- A liquid continuum {r-l} was created by varying F3: for the [r]-endpoint, it fell to 2000 Hz, and for the [l]-endpoints, it rose to 2800 Hz.
- The continuum was created with 6 step increments.
- The liquid portion was followed by a 95 ms gap with low-frequency periodic energy to mimic closure voicing of [d] and [g].
- The [d]-[g] continuum was created by varying F3: in the [da] endpoint, F3 began at 2690 Hz, while in the [ga] endpoint, it began at 2104 Hz, again with 6 step increments.

# Illustration with spectrograms

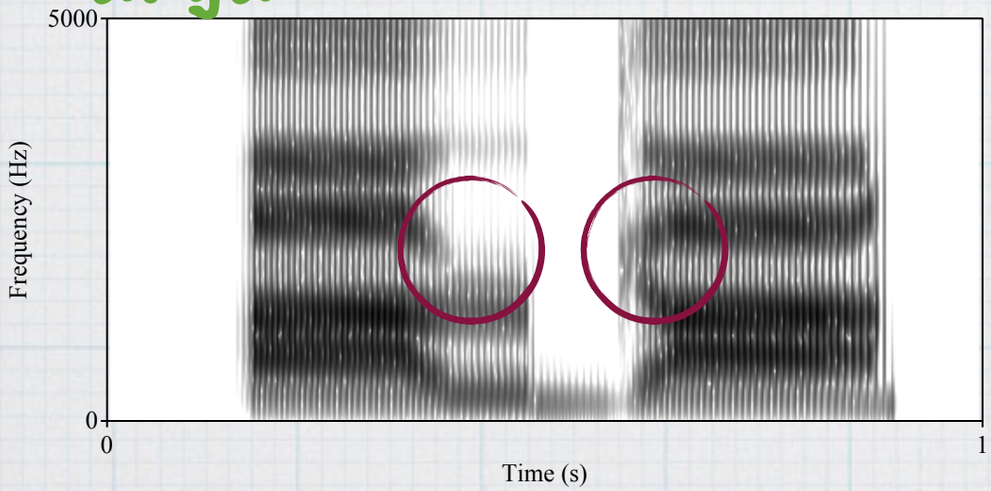
## arda



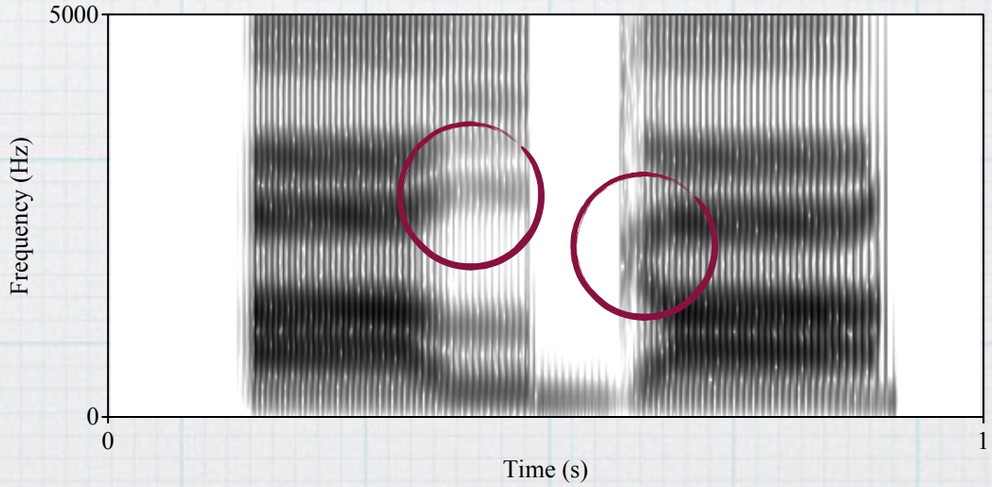

## alda




## arga




## alga



# Procedure

- In the listening phase, listeners heard one stimulus and were asked to judge whether the second syllable was [da] or [ga].
- The order of the stimuli was randomized within each block.
- All listeners went through 8 blocks.
- In the second phase of the experiment, the listeners were presented with the [ar] and [al] endpoint stimuli in isolation, and were asked to identify these sounds (20 trials). D-prime was calculated for each listener as a measure of their ability to perceive the difference between [r] and [l].
- 30 native speakers of Japanese participated in this study.

# Analyzing the identification patterns

* Various logistic models were fit, and the model with the best AIC was chosen.

* $\text{logit}(Y) = \beta_0 + \beta_1 DG + \beta_2 RL + \beta_3 DG*RL + e$

* $\beta_2$ = context effect

# D-prime

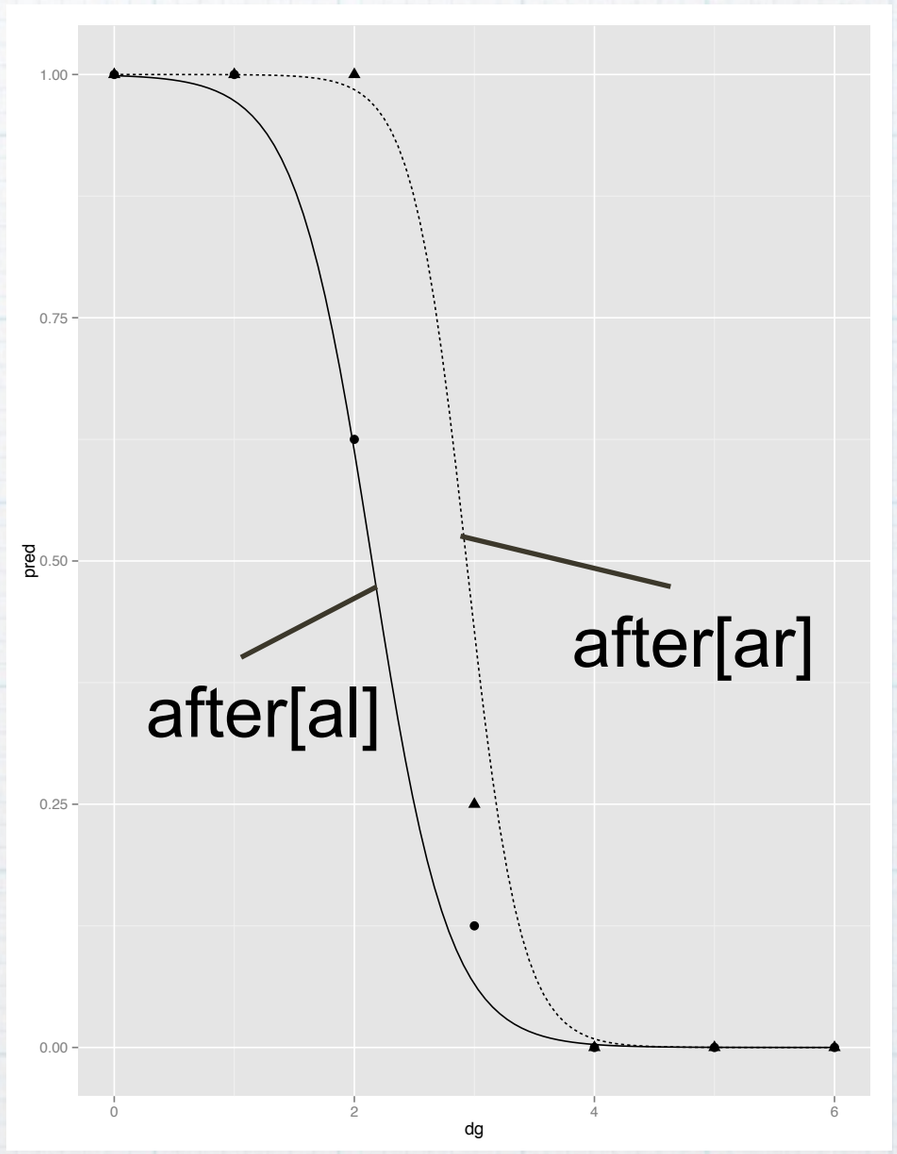
| stimuli<br>response | l (True) | r (False) |
|---|---|---|
| l (True) | hit | false alarm |
| r (False) | miss | correct rejection |

d-prime = z(hit)-z(FA)

Measure of the ability to distinguish /r/ and /l/. Higher d-prime values indicate higher sensitivity to the contrast.

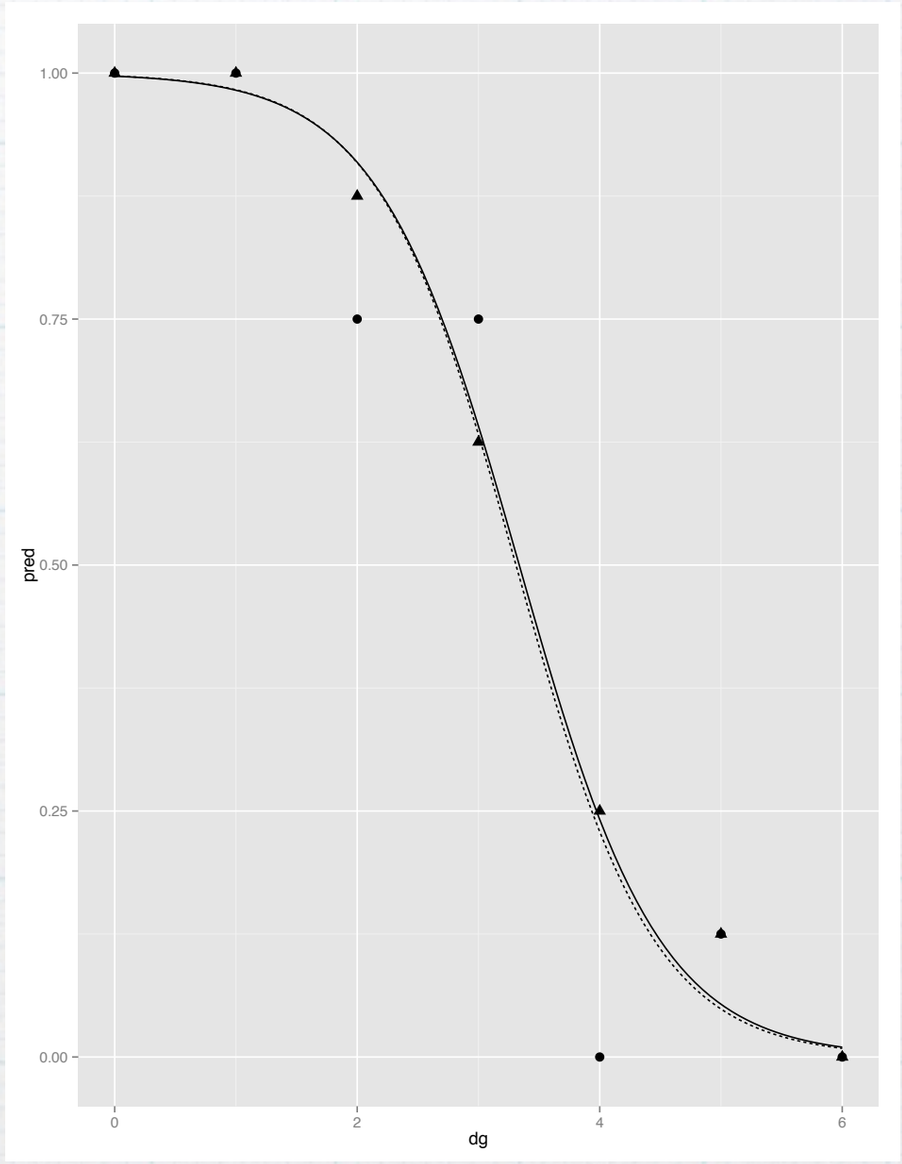
# Prediction 1

Those who are sensitive to the [l]-[r] distinction would show strong context effect.



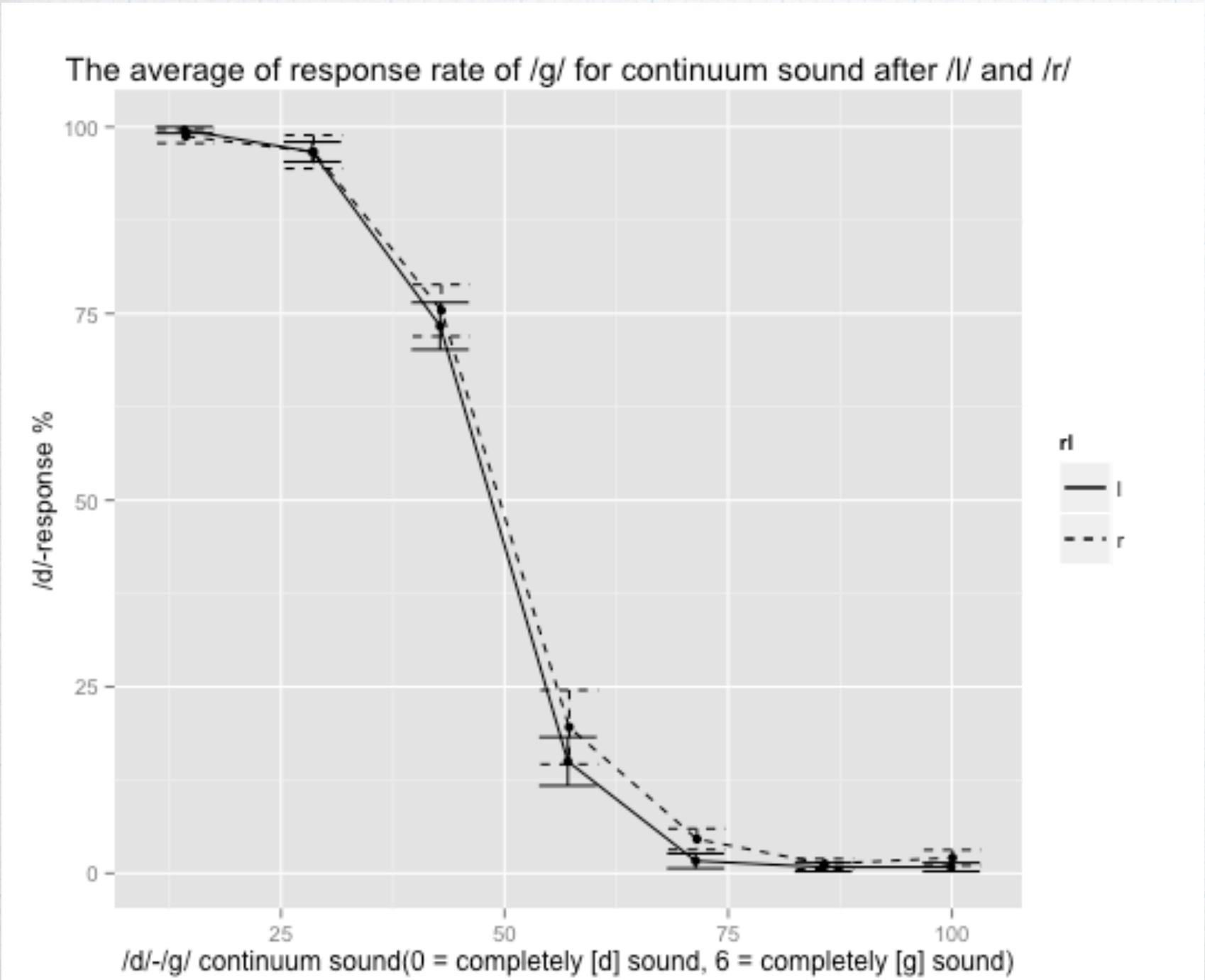
# Prediction 2

Those who are not sensitive to the [r]-[l] distinction would show weak context effect.



# Results

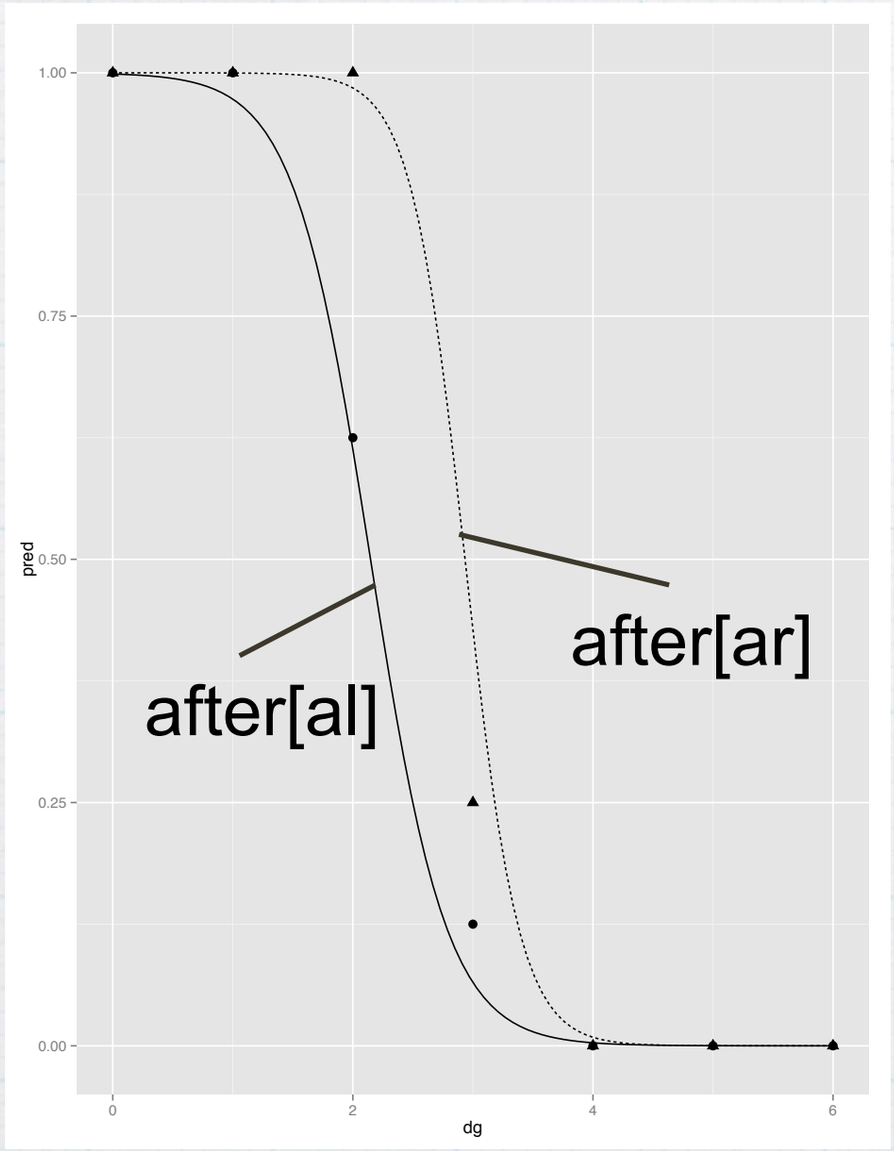
# Average identification functions



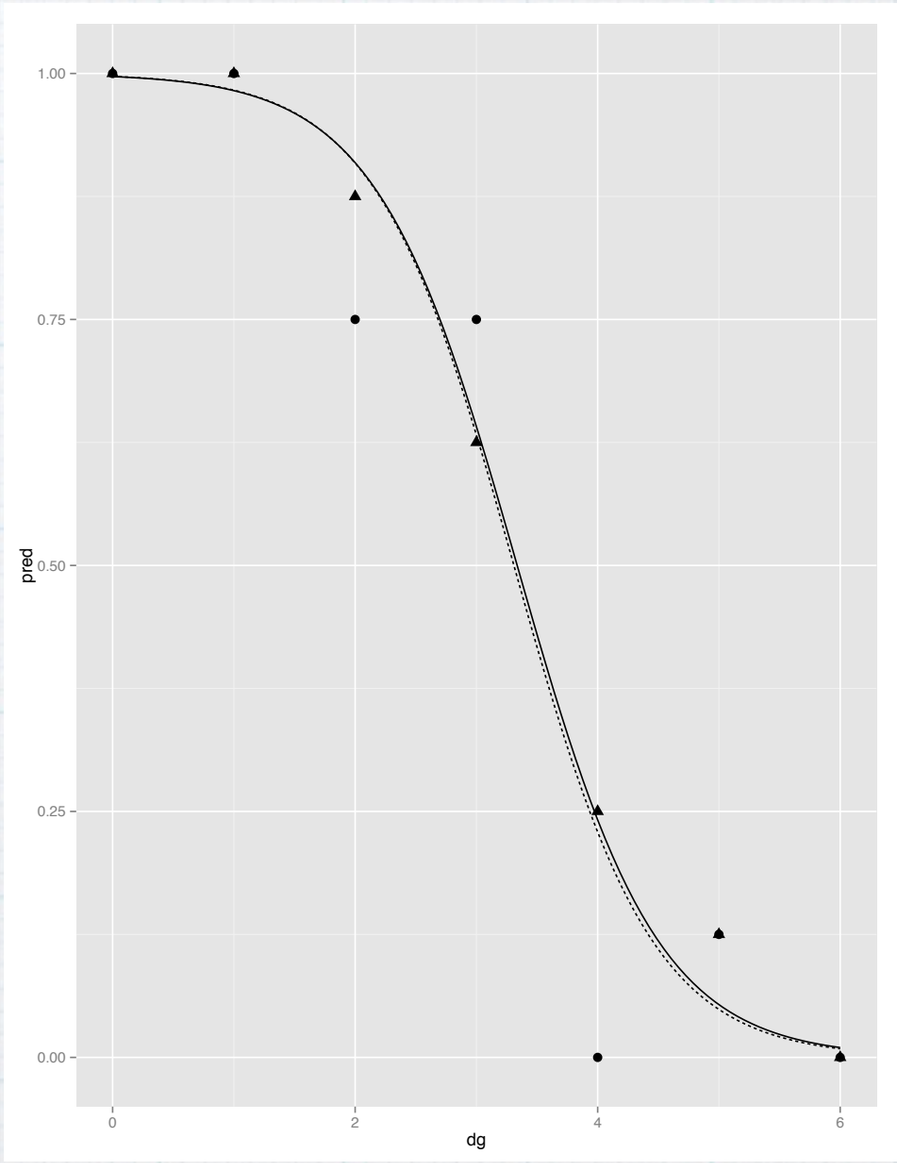

β<.001

No context effects for Japanese listeners?

# Interspeaker differences




β=1.05

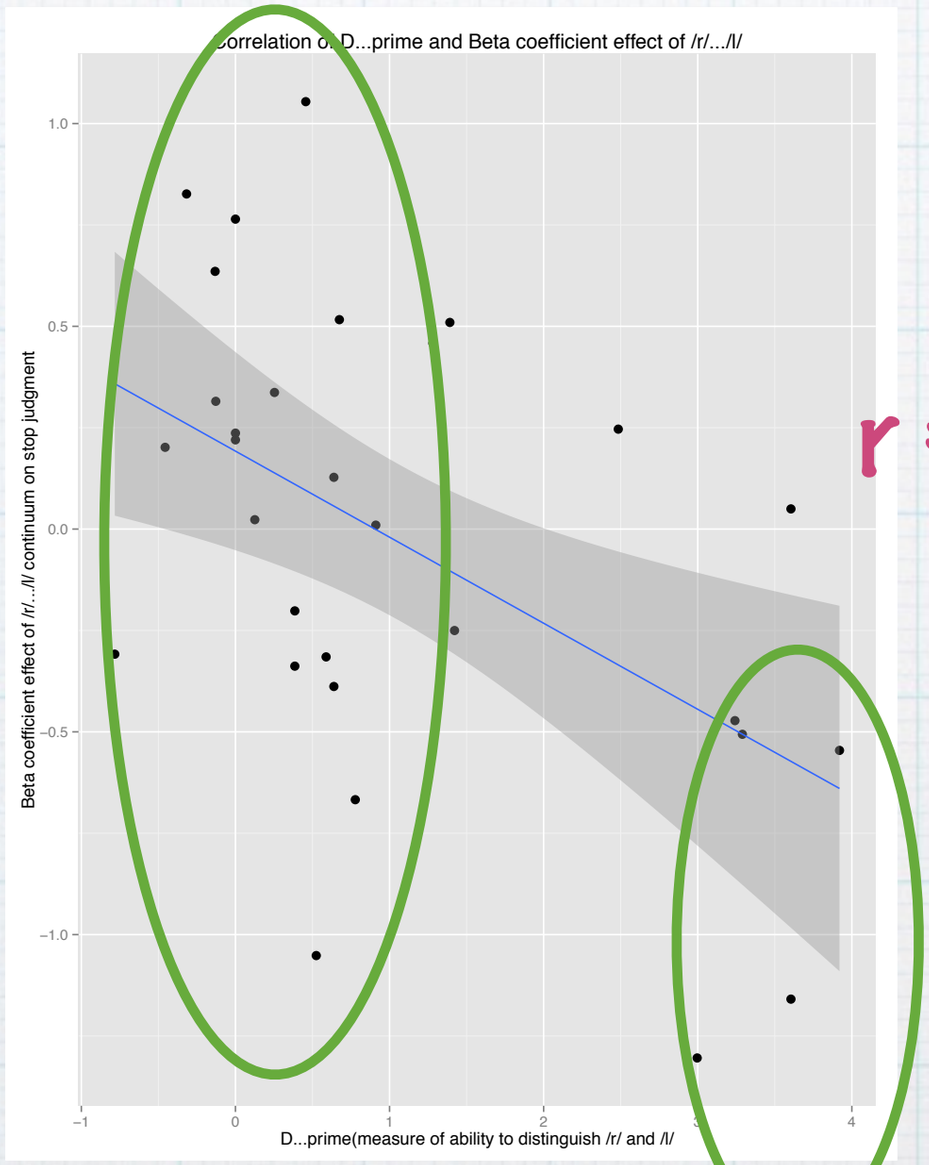



β<.001

# Correlation with d' and magnitude of context effect

Those with low d-prime values can differ in how they are affected by context effect.

Magnitude of context effect

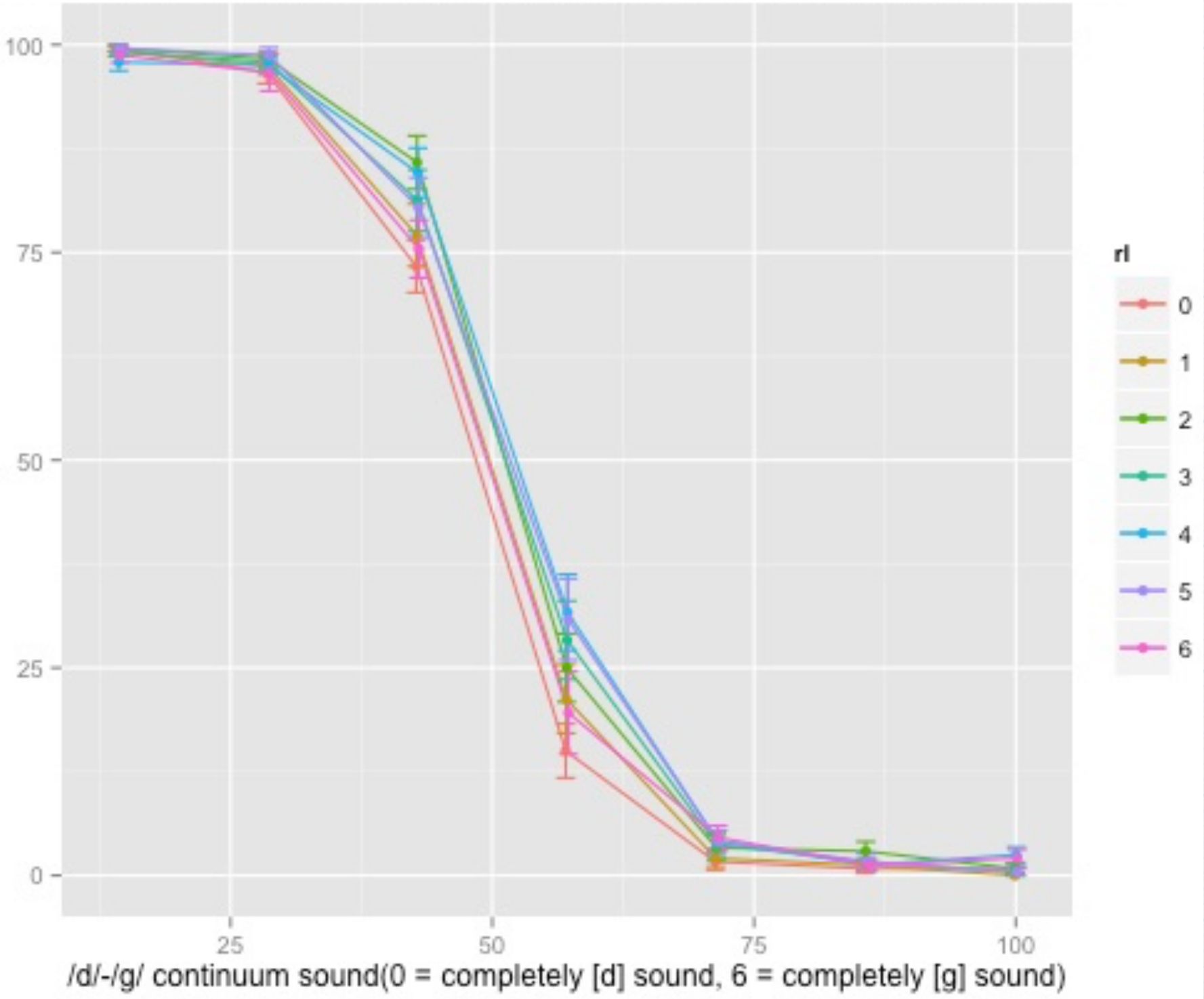



r = -0.4, p <.01

Those with high d-prime values show "anti-compensation for compensation" effect

Ability to distinguish [l] and [r]

# All the data together




It is not the case that the most [r]-like liquid induces the most [d] responses.

This is not expected from the auditory contrast theory, perhaps hard to explain in the compensation for coarticulation theory

# Result summary

* There are three groups of Japanese listeners:
    1. who show expected context effect.
    2. who show unexpected context effect (i.e. assimilator).
    3. who are insensitive.
* Those who can distinguish [r] and [l] tend to belong to Group 2.
* The relationship between the liquid's F3 and the perceived F3 of the following stop is not (negatively) linear.

# What do the current results say about the theories of speech perception?

* These results are predicted by neither the compensation for coarticulation theory or general auditory contrast.

* We could only partially replicate Mann (1986).

* After all, where does "assimilation effect" come from?

* "Mis-parsing" explanation pursued in Kingston's lab at UMass; e.g. low frequency of [r]'s F3 is "mis-parsed" as information belonging to the stop, inducing [g]-responses.

* But why does mis-parsing happen and when?

# Why assimilation?

* Those who know English well may be sensitive to lexical statistics.

* The IPhOD calculator (Vaden et al 2009):

| | | | | |
|---|---|---|---|---|
| rd | 0.00380 | ld | 0.00244 | raw frequency |
| rg | 0.00068 | lg | 0.00011 | |
| rd | 0.848 | ld | 0.957 | conditional probability |
| rg | 0.152 | lg | 0.043 | |

* Bias toward [d] is slightly stronger after [r] than after [l].
* No explicit instructions that the stimuli were English words.

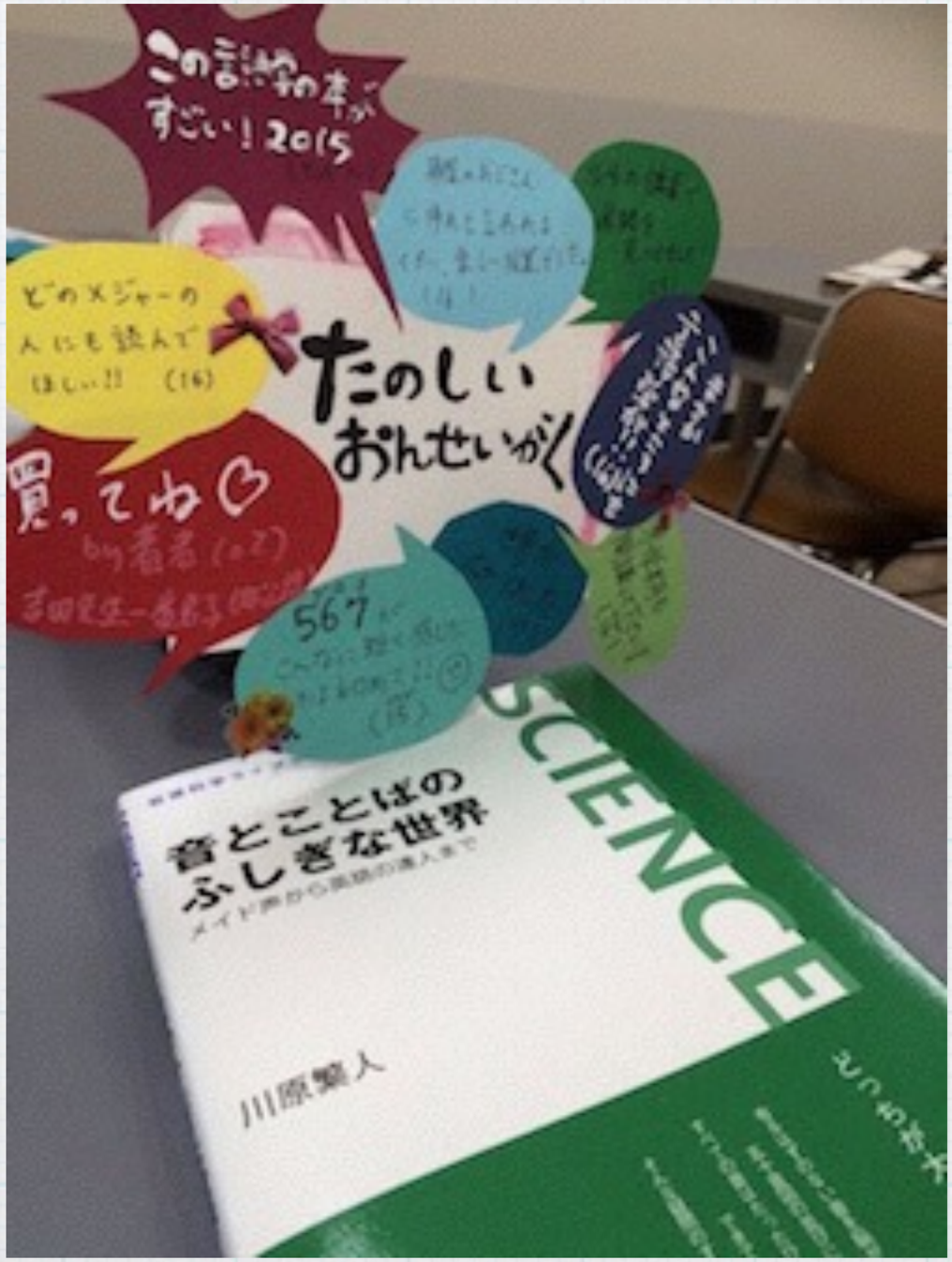
# Discussion and remaining questions

* Not all Japanese speakers show context effect due to [l] and [r].

* The results are not compatible with either compensation for coarticulation or general auditory contrast.

* What's the mechanism behind "assimilation"?

# My teacher told me to...



# Acknowledgments


* John Kingston for our collaboration (past and present)
* JSPS grants to the second author: #26770147 and #26284059.
* Members of the Keio phonetics/phonology study group.
* Participants at Tokyo Circle of Participants (1/30/2016).